

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



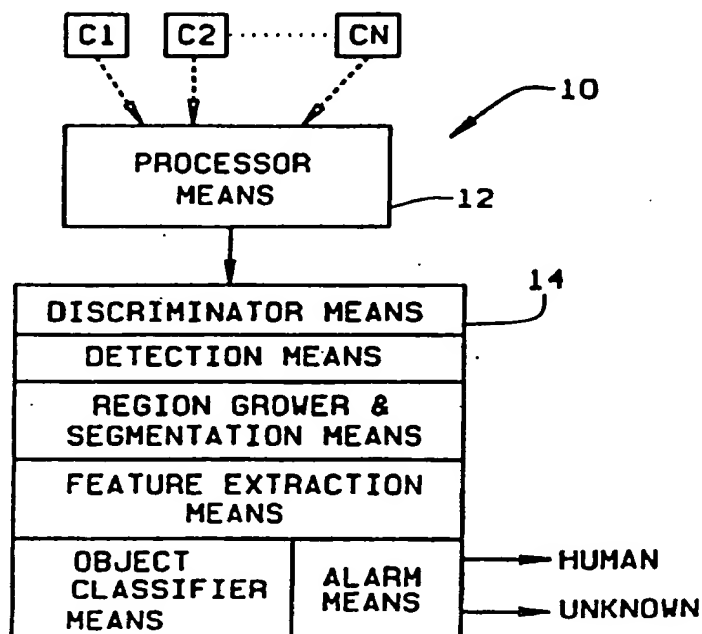
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>G06K 9/20, 9/46, G08B 23/00, H04N 7/18</b>		<b>A1</b>	(11) International Publication Number: <b>WO 98/28706</b>
			(43) International Publication Date: 2 July 1998 (02.07.98)
(21) International Application Number: <b>PCT/US97/24163</b>		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 23 December 1997 (23.12.97)			
(30) Priority Data: 08/772,595 23 December 1996 (23.12.96) US 08/771,991 23 December 1996 (23.12.96) US 08/772,731 23 December 1996 (23.12.96) US			
(71) Applicant: ESCO ELECTRONICS CORPORATION [US/US]; Suite 200, 8888 Ladue Road, St. Louis, MO 63124 (US).			
(72) Inventors: WOOTTON, John, R.; 700 Rugby Court, St. Louis, MO 63141 (US). WALDMAN, Gary, S.; 12934 Walnut Way, St. Louis, MO 63146 (US). HOBSON, Gregory, L.; 17 Upper Dardenne Farms Drive, St. Charles, MO 63304 (US).			
(74) Agent: MULLER, J., Joseph; Polster, Lieder, Woodruff & Lucchesi, 763 South New Ballas Road, St. Louis, MO 63141 (US).			
		<p><b>Published</b></p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	

(54) Title: LOW FALSE ALARM RATE VIDEO SECURITY SYSTEM USING OBJECT CLASSIFICATION

(57) Abstract

A video detection system (10) and method detects an intruder from video images of a scene. The method employs a recognition process to differentiate between humans and animals. The method is used only after possible false alarms resulting from identified effects of noise, aliasing, non-intruder motion occurring within the scene, and effects of global or local lighting changes. The object recognition process includes determining the regions containing a potential intruder, outlining and growing those regions to encompass all of potential intruders, determining a set of shape features from the region and eliminating possible shadow effects, normalizing the set, and comparing the normalized set with sets of features of humans and animals. This comparison produces a confidence level indicating a human intruder. An alarm is given for a sufficiently high confidence level. The possibility of a false alarm due to an animal or a non-identifiable object is also substantially eliminated.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## LOW FALSE ALARM RATE VIDEO SECURITY SYSTEM USING OBJECT CLASSIFICATION

### Technical Field

5           This invention relates to video security systems and a method for detecting the presence of an intruder into an area being monitored by the system; and more particularly, to i) the rejection of false alarms which might otherwise occur because of global or local, natural or manmade, lighting changes which occur within a scene observed by the system, ii) the discernment  
10 of an intruder based upon sensed surface differences which occur within the scene rather than lighting changes which may occur therewithin, and iii) the classification of an intruder detected by the system as either human or non-human, and to provide an alarm if the intruder is classified as a human.

### Background Art

15           A security system of the invention uses a video camera as the principal sensor and processes a resulting image to determine the presence or non-presence of an intruder. The fundamental process is to establish a reference scene known, or assumed, to have no intruder(s) present. An image of the present scene, as provided by the video camera, is compared with an image of  
20 the reference scene and any differences between the two scenes are ascertained.

          If the contents of the two scenes are markedly different, the interpretation is that an intrusion of some kind has occurred within the scene. Once the possibility of an intrusion is evident, the system and method operate to first eliminate possible sources of false alarms, and to then classify any remaining  
25 differences as being the result of a human or non-human intrusion. Only if a determination is made that the anomaly results from a human intrusion is a notification (alarm) made. All other anomalies which produce a difference between the two images are identified as false alarms for which no notification is given.

30           One issue addressed in making the determination is the possibility of false alarms caused by lighting changes within a scene, whether natural or

manmade, global or local. As discussed therein, the differences between the reference scene and a later scene resulting from lighting effects can be now be identified so that no false alarm results from them. However, there are other potential causes of false alarms which also must be recognized. The video security system and image processing methodology as described herein recognizes anomalies resulting from these other causes so these, too, can be accounted for. The method includes comparing, on a pixel by pixel basis, the current image with the reference image to obtain a difference image. Any non-zero pixel in the difference image indicates the possible presence of an intrusion, after image artifacts such as noise, aliasing of the video, and movement within the scene not attributable to a life form (animal or human) such as the hands of a clock, screen savers on computers, oscillating fans, etc., have been accounted for. Because the system and method use an absolute difference technique with pixel by pixel subtraction, the process is sensitive to surface differences between the scene but insensitive to light on dark or dark on light changes, and thus is very sensitive to any intrusion within the scene. Furthermore, each pixel represents a gray level measure of the scene intensity that is reflected from that part of the scene. The gray level intensity can alter for a variety of reasons, the most relevant of these being that there is a new physical presence at that particular part of the scene.

Two important features of the video security system is to inform an operator/verifier of the presence of a human intruder, and to not generate false alarms. Thus to be economically viable, and not place unduly high demands on the operator/verifier, the system must operate to eliminate as many false alarms as possible without impacting the overall probability of detecting an intruder's presence. A fundamental cause of false alarms stems from the sensor and methodology used to ascertain if an intrusion has occurred. By use of the processing methodology described herein, various effects which could otherwise trigger false alarms are accounted for so that only a life form intruding into the scene will produce an alarm. However, even though

unwanted detections due to non animal/non-human caused motion are eliminated, it is still necessary to differentiate between a class of human motion and a class of non-human or animal motion. Only by doing so can intrusions resulting from human actions properly cause an alarm to be given and false  
5 alarms resulting from animal movements not be provided.

Previous attempts have made to provide a reliable security system. These systems have relied upon contact break mechanisms or PID (passive infra red) motion sensors to detect intruder presence. Examples of the use of infrared devices, either as a passive element or as a scanning device, are disclosed in U.  
10 S. patents 5,283,551, 5,101,194, 4,967,183, 4,952,911, 4,949,074, 4,939,359, 4,903,990, 4,847,485, 4,364,030, and 4,342,987. More recently, however, the realization that an image processor is required to transmit the video for confirmation purposes has led to the development of using the image processor to actually detect the possible presence of an intruder. Such a system has an  
15 economy of hardware and obviates the need for PID sensors or contact breaker devices. A security system of this type has comparable performance to a PID counterpart. However, there are areas where considerable benefits accrue if false alarms, which occur due to the erroneous injection of light into the scene without the presence of an intruder, are reduced or eliminated.

20 The cause of these false alarms stem from the sensor and methodology used to ascertain if an intrusion has occurred. As stated earlier, a past image of the scene being surveyed is compared with the present scene as taken from the camera. The form of comparison is essentially a subtraction of the two scenes on a pixel by pixel basis. Each pixel represents a gray level measure of the  
25 scene intensity that is reflected from that part of the scene. Gray level intensity can change for a variety of reasons, the most important being a new physical presence within a particular part of the scene. Additionally, the intensity will change at that location if the overall lighting of the total scene changes (a global change), or the lighting at this particular part of the scene changes (a local  
30 change), or the AGC (automatic gain control) of the camera changes, or the

ALC (automatic light level) of the camera changes. With respect to global or local lighting changes, these can result from natural lighting changes or manmade lighting changes. Finally, there will be a difference of gray level intensity at a pixel level if there is noise present in the video. Only the situation  
5 of a physical presence in the scene is a true alarm; the remainder all comprise false alarms within the system. For a security system to be economically viable and avoid an unduly high load on an operator who has to verify each alarm, the system must process images in a manner which eliminates as many of false alarms as possible without impacting the overall probability of detecting the  
10 presence of an intruder.

Some efforts have previously been made in attempting to recognize objects, including humans, whose presence is detected or sensed in an image. For example, U.S. patent 5,305,390 to Frey et al., teaches automatic recognition and classification of persons or objects as they pass through a doorway or  
15 entrance. The intrinsic sensor is an active laser beam, and the system of Frey et al. operates by measuring the height of an object passing through an aperture (doorway) to classify the object as a person or not. Therefore, the system is a height discriminator rather than an object recognition or classification system. Thus, for example, if a person crawls through the aperture, they will probably be  
20 designated as a non-human.

U.S. patent 5,289,275 to Ishii et al., is directed to a surveillance monitoring system using image processing for monitoring fires and thefts. The patent teaches use of a color camera for monitoring fires and a method of  
25 comparing the color ratio at each pixel in an image to estimate the radiant energy represented by each pixel. A resulting ratio is compared to a threshold with the presence of a fire being indicated if the threshold is surpassed. A similar technique for detecting the presence of humans is also described. The patent teaches the use of image processing together with a camera to detect the presence of fires and abnormal objects.

30 U.S. patent 4,697,097 to Yausa et al. also teaches use of a camera to

detect the presence of an object. Once an anomaly is detected because of differences in the comparison of an original and a later image, the system automatically dials and sends a difference image, provided the differences are large enough, to a remote site over a telephone line. At the remote site, the  
5 image is viewed by a human. While teaching some aspects of detection, Yausa et al. does not go beyond the detection process to attempt and use image processing to recognize that the anomaly is caused by a human presence.

U.S. patent 4,257,063, which is directed to a video monitoring system and method, teaches that a video line from a camera can be compared to the  
10 same video line viewed at an earlier time to detect the presence of a human. However, the detection device is not a whole image device, nor does it make any compensation for light changes, nor does it teach attempting to automatically recognize the contents of an image as being derived from a human. Similarly, U.S. patent 4,161,750 teaches that changes in the average  
15 value of a video line can be used to detect the presence of an anomalous object. Whereas the implementation is different from the '063 patent, the teaching is basically the same.

All of these previous attempts at recognition have certain drawbacks, whether the type of imaging, method of processing, etc., which would result in  
20 either an alarm not being provided when one should, or in false alarms being given. The system and method of the present invention overcome these problems or shortcomings to reliably provide accurate indications of human intrusion in an area being monitored by a security system. Such an approach is particularly cost efficient because it reduces the necessity of guards having to  
25 patrol secured areas (which means each area will be observed only on an infrequent basis unless there are a large number of guards), while ensuring that any intrusion in any area is not only observed, but an appropriate alarm is sounded in the event of a human intrusion.

#### Disclosure of the Invention

30 Among the several objects of the present invention may be noted the

provision of a video security system and method for visually monitoring a scene and detecting the presence of an intruder within the scene;

the provision of such a system and method whose operation is based upon the premise that only the presence of a human intruder is of consequence  
5 to the security system, with everything else constituting a false alarm;

the provision of such a system and method to readily distinguish between changes within the scene caused by the presence of a person entering the scene as opposed to changes within the scene resulting from lighting changes (whether global or local, natural or man made) and other anomalies which occur within the  
10 scene to detect the presence of an intruder;

the provision of such a system and method to employ a recognition process rather than an abnormality process such as used in other systems to differentiate between human and non-human objects, so to reduce or substantially eliminate false alarms;

15 the provision of such a system and method to provide a high probability of detection of the presence of a human, while having a low probability of false alarms;

the provision of such a system and method which provides image processing such that false alarms resulting from the inadvertent presence of  
20 artifacts as caused by noise, aliasing, non-intruder motion occurring within the scene, are identified and do not provoke a system response;

the provision of such a system and method which, once an intruder has been detected, further classifies the intrusion as resulting from the presence of a human life form, or the presence of non-human life forms, such as shadows,  
25 dogs, cats, rats, mice, birds, etc.

the provision of such a system and method in which an indication of an intrusion is given only after the cause of the intrusion has been determined as resulting from the presence of a human so to avoid giving false alarms;

the provision of such a system and method to also provide a second and  
30 lower level alarm in the event an object cannot be classified as human or non-



human so an operator/verifier is informed of the possible presence of an intruder in the scene;

the provision of such a system and method to evaluate a series of images of the scene and determine, for each image examined, the classification of an object so to have an increased confidence level that an object classified as a  
5 human is properly classified;

the provision of such a system and method in which the alarm indication which is provided includes automatically accessing a site remote from the scene where the intrusion occurs and transmitting an image of the scene in which the  
10 intruder is present to the remote site;

the provision of such a system and method in which the transmitted image is a compressed image of the scene rather than a small subset of the image; and,

the provision of such a system and method by which a number of areas  
15 can be continuously, reliably, and cost effectively monitored with a human intrusion in any area being reliably detected and the appropriate alarm given.

In accordance with the invention, generally stated, a video detection system detects the presence of an intruder in a scene from video provided by a camera observing the scene. A recognition process differentiates between  
20 human and non-human (animal) life forms. The presence of a human is determined with a high degree of confidence so there is a very low probability of false alarms. Possible false alarms resulting from the effects noise, aliasing, non-intruder motion occurring within the scene, and the effects of global or local lighting are first identified and only then is object recognition performed.

25 Performing object recognition includes determining which regions within the image may be an intruder, outlining and growing those regions so the result encompasses all of what may be the intruder, determining a set of shape features from the region and eliminating possible shadow effects, normalizing the set of features and comparing the resulting set with sets of features for humans and non-  
30 human (animal) life forms. The result of the comparison produces a confidence

level as to whether or not the intruder is a human. If the confidence level is sufficiently high, an alarm is given. By performing object classification in this manner, the further possibility a false alarm may occur due to the presence of an animal, or a non-identifiable object in the scene is also substantially eliminated.

5 Other objects and features will be in part apparent and in part pointed out hereinafter.

#### Brief Description of Drawings

In the drawings, Fig. 1 is a simplified block diagram of a video security system of the present invention for viewing a scene and determining the  
10 presence of an intruder in the scene;

Fig. 2 is a representation of an actual scene viewed by a camera of the system;

Fig. 3 is the same scene as Fig. 2 but with the presence of an intruder;

Fig. 4 is a representation of another actual scene under one lighting  
15 condition;

Fig. 5 is a representation of the same scene under different lighting conditions and with no intruder in the scene;

Fig. 6A is a representation of the object in Fig. 3 including its shadow, Fig. 6B illustrates outlining and segmentation of the object; and Fig. 6C  
20 illustrates the object with its shadow removed and as resampled for determining a set of features for the object;

Figs. 7A-7C represent non-human (animal) life forms with which features of the object are compared to determine if the object represents a human or non-human life form and wherein Fig. 7A represents a cat, Fig. 7B a  
25 dog, and Fig. 7C a bird;

Figure 8 is a simplified time line indicating intervals at which images of the scene are viewed by the camera system;

Figure 9 represents a pixel array such as forms a portion of an image; and,

30 Fig. 10 illustrates masking of an image for those areas within a scene

where fixed objects having an associated movement or lighting change are located.

Corresponding reference characters indicate corresponding parts throughout the drawings.

5 Best Mode for Carrying Out the Invention

Referring to the drawings, a video security system of the invention is indicated generally 10 in Fig. 1. The system employs one or more cameras C1-Cn each of which continually views a respective scene and produces a signal representative of the scene. The cameras may operate in the visual or infrared portions of the light spectrum and a video output signal of each camera is supplied to a processor means 12. Means 12 processes each received signal from a camera to produce an image represented by the signal and compares the image representing the scene at one point in time with a similar image of the scene at a previous point in time. The signal from the imaging means represented by the cameras may be either an analog or digital signal, and processing means 12 may be an analog, digital, or hybrid processor.

In Fig. 2, an image of a scene is shown, the representation being the actual image produced by a camera C. Fig. 2 represents, for example, a reference image of the scene. Fig. 3 is an image exactly the same as that in Fig. 2 except that now a person (human intruder) has been introduced into the scene. Fig. 3 is again an actual image produced by a camera C. Similarly, Fig. 4 represents a reference image of a scene, and Fig. 5 a later image in which there is a lighting change but not an intrusion. The system and method of the invention operate to identify the presence of such a human intruder and provide an appropriate alarm. However, it is also a principal feature of the invention to not produce false alarms. As described herein and in the referenced co-pending application, there are numerous sources of false alarms and using a series of algorithms employed by the invention, these sources are identified for what they are so no false alarms are given.

30 Operation of the invention is such that segments of an image (Fig. 3, Fig.

5) which differ from segments of an earlier image (Fig. 2, Fig. 4) are identified. A discriminator means 14 evaluates those segments to determine if the differences are caused by a local lighting change within the scene (Fig. 5), or the movement of an intruder within the scene (Fig. 3). As noted, if the change is caused by an intruder, an alarm is given. But, if the differences result from global or local lighting changes, the effects of motion of objects established within the scene, noise, and aliasing effects, these are recognized as such so false alarm is not given. Detection of local lighting changes such as shown in Fig. 5 are described in the referenced co-pending application.

10 Generally, a single processor can handle several cameras positioned at different locations within a protected site. In use, the processor cycles through the different cameras, visiting each at a predetermined interval. At system power-up, the processor cycles through all of the cameras doing a self-test on each. One important test at this time is to record a reference frame against  
15 which later frames will be compared. A histogram of pixel values is formed from this reference frame. If the histogram is too narrow, a message is sent to the effect that this camera is obscured and will not used. This is done to guard against the possibility of someone obscuring the camera while it is off by physically blocking the lens with an object or by spray-painting it. If a camera  
20 is so obscured, then all the pixel values will be very nearly the same and this will show up in the histogram. Although the camera is now prevented from participating in the security system, the system operator is informed that something is amiss at that particular location so the problem can be investigated.

In accordance with the method, a reference frame f1 is created.  
25 Throughout the monitoring operation, this reference frame is continuously updated if there is no perceived motion within the latest image against which a reference image is compared. At each subsequent visit to the camera a new frame f2 is produced and subtracted from the reference. If the difference is not significant, the system goes on to the next camera. However, if there is a  
30 difference, frame f2 is stored and a third frame f3 is created on the next visit and

compared to both frames f1 and f2. Only if there is a significant difference between frames f3 and f2 and also frames f3 and f1, is further processing done.

This three frame procedure eliminates false alarms resulting from sudden, global light changes such as caused by lightning flashes or interior lights going

5 on or off. A lightning flash occurring during frame f2 will be gone by frame f3, so there will be no significant difference between frame f3 and f1. On the other hand, if the interior lights have simply gone on or off between frames f1 and f2, there will be no significant changes between frames f2 and f3. In either instance, the system proceeds on to the next camera with no more processing.

10 Significant differences between frames f1 and f2, frames f3 and f2, and frames f3 and f1 indicate a possible intrusion requiring more processing.

Besides global lighting changes occurring between the images, non-intruder motion occurring within the scene is also identified so as not to trigger processing or cause false alarms. Thus, for example, if the fan shown in the lower

15 left portion of Figs. 4 and 5 were running, movement of the fan blades would also appear as a change from one image to another. Similarly, if the fan is an oscillating fan, its sweeping movement would also be detected as a difference from one image to another. As described hereinafter, and as shown in Fig. 10, the area within the scene where an object having an associated movement is generally

20 fixed and its movement is spatially constrained movement, the area where this movement occurs is identified and masked so, in most instances, motion effects resulting from operation of the object (fan) are disregarded. Although, if the motion of an intruder overlaps the masked area, the difference from one image to another is identified and further processing, including the normally masked area

25 takes place. It will be understood that there are a variety of such sources of apparent motion which are identified and masked. Besides the fan, there are clocks both digital and those having hands. In one instance, the numerical display of time changes; in the other instance, the hands of the clock (particularly the second hand) has a noticeable movement. Computers with screen savers may

30 have a constantly changing image on their monitors. In manufacturing areas,

different pieces of equipment, rotating or reciprocal machinery, robotic arms, etc., all exhibit movements which can be identified and accounted for during processing.

Any video alert system which uses frame-to-frame changes in the video to detect intrusions into a secured area is also vulnerable to false alarms from the inadvertent (passing automobile lights, etc.) or deliberate (police or security guard flashlights) introduction of light into the area, even though no one has physically entered the area. The system and method of the invention differentiate between a change in a video frame due to a change in the irradiation of the surfaces in the FOV (field of view) as in Fig. 5, and a change due to the introduction of a new reflecting surface in the FOV as in Fig. 3. The former is then rejected as a light "intrusion" requiring no alarm, whereas the latter is identified as a human intruder for which an alarm is given. It is important to remember that only the presence of a human intruder is of consequence to the security system, everything else constitutes a false alarm. It is the capability of the system and method of the invention to yield a high probability of detection of the presence of a human, while having a low probability of false alarms which constitutes a technically differentiated video security system. The video processing means of the present invention can also defeat the artifacts of noise, aliasing, screen savers, oscillating fans, drapery blown by air flow through vents, etc.

#### ALGORITHM PROCESS STEPS

The complete algorithm processes that are implemented by the method of the present invention are as follows:

- 25     Antialiasing;
- Detection (Differencing and Thresholding)
- Outlining;
- Region Grower Segmentation;
- Noise removal;
- 30     Shadow removal;

- 13 -

Tests for global and local lighting changes;

Masking;

Shape features;

Fourier Descriptors;

5       Object classification

#### ANTIALIASING PROCESS

The alias process is caused by sampling at or near the intrinsic resolution of the system. As the system is sampled at or near the Nyquist frequency, the video, on a frame by frame basis, appears to scintillate, and certain areas will  
10       produce Moire like effects. Subtraction on a frame by frame basis would cause multiple detections on scenes that are unchanging. In many applications where this occurs it is not economically possible to over sample. Elimination of aliasing effects is accomplished by convolving the image with an equivalent two-dimensional (2D) smoothing filter. Whether this is a  $3 \times 3$  or  $5 \times 5$  filter, or  
15       a higher filter, is a matter of preference as are the weights of the filter.

#### DETECTION PROCESS

The detection process consists of comparing the current image to a reference image. To initialize the system it is assumed that the operator has control over the scene and, therefore, will select a single frame for the reference  
20       when there is nothing present. (If necessary, up to 60 successive frames can be selected and integrated together to obtain an averaged reference image). As shown in Fig. 1, apparatus 10 employs multiple cameras C1-Cn, but the methodology with respect to one camera is applicable for all cameras. For each camera, an image is periodically selected and the absolute difference between  
25       the current image (suitably convolved with the antialiasing filter) and the reference is determined. The difference image is then thresholded (an intensity threshold) and all of the pixels exceeding the threshold are accumulated. This step eliminates a significant number of pixels that otherwise would result in a non-zero result simply by differencing the two images. Making this threshold  
30       value adaptive within a given range of threshold values ensures consistent

performance. If the count of the pixels exceeding the intensity threshold exceeds a pixel count threshold, then a potential detection has occurred. At this time, all connected hit pixels (pixels that exceed the intensity threshold) are segmented, and a count of each segmented object is taken. If the pixel count of  
5 any object exceeds another pixel count threshold, then a detection is declared. Accordingly, detection is defined as the total number of hit pixels in the absolute difference image being large and there is a large connected object in the absolute difference image.

With respect to noise, the key to rejecting noise induced artifacts is their  
10 size. Noise induced detections are generally spatially small and distributed randomly throughout the image. The basis for removing these events is to ascertain the size (area) of connected pixels that exceed the threshold set for detection. To achieve this, the region where the detected pixels occur is grown into connected "blobs". This is done by region growing the blobs. After region  
15 growing, those blobs that are smaller in size than a given size threshold are removed as false alarms.

#### REGION GROWER SEGMENTATION

Typically, a region growing algorithm starts with a search for the first object pixel as the outlining algorithm does. Since searching and outlining has  
20 already been performed, and since the outline pixels are part of the segmented object, these do not need to be region grown again. Outline pixel arrays are now placed on a stack, and the outline pixels are zeroed out in the absolute difference image. A pixel is then selected (removed from the stack) and the outline pixels are zeroed out in the absolute difference image. The selected  
25 pixel P and all of its eight neighbors P1-P8 (see Fig. 9) are examined to see if hit points occur (i.e. they are non-zero). If a neighbor pixel is non-zero, then it is added to the stack and zeroed out in the absolute difference image. Note that for region growing, all eight neighboring pixels are examined, whereas in outlining, the examination of neighboring pixels stops as soon as an edge pixel  
30 is found. Thus, in outlining, as few as one neighbor may be investigated. The



region growing segmentation process stops once the stack is empty.

One way to achieve the desired discrimination is to use an elaboration of the retinex theory introduced by Edwin Land some 25 years ago. Land's theory was introduced to explain why human observers are readily able to identify  
5 differences in surface lightness despite greatly varying illumination across a scene. Although the following discussion is with regards to a human observer, it will be understood that besides human vision, Land's theory is also applicable to viewing systems which function in place of a human viewer. According to the theory, even if the amount of energy reflected (incident energy times surface  
10 reflectance) from two different surfaces is the same, an observer can detect differences in the two surface lightness' if such a difference exists. In other words, the human visual system has a remarkable ability to see surface differences and ignore lighting differences. Land's hypothesis was that this ability derives from comparison of received energies across boundaries in the  
15 scene. Right at any boundary, light gradients make no difference because the energies received from adjacent regions on opposite sides of a boundary are in the correct ratio (the same as the ratio of reflectances). Furthermore, correct judgments about lightness' of widely separated regions are made by a serial process of comparisons across intervening regions. At first the theory was  
20 applied only to black and white scenes. Subsequently, it was extended to color vision by assuming that three separate retinex systems judge the lightness of surfaces in the three primary colors (red, green and blue). The retinex theory of color vision is able to explain why surface colors appear very stable to humans even though the nature of the illumination may change through a wide range.

25 It is the ability to discern surface differences and ignore lighting changes which is incorporated into the video security system and method of the present invention. Therefore, whether or not Land's theory correctly explains the way human vision operates, use of his concepts in the present invention make the system and method immune to light "intrusions".

30 A video signal (gray level) for any pixel is given by

$$g \propto \int E(\lambda) r(\lambda) S(\lambda) d\lambda \quad (1)$$

where  $E(\lambda) \equiv$  scene spectral irradiance at the pixel in question  
 $r(\lambda) \equiv$  scene spectral reflectance at the pixel in question  
 $S(\lambda) \equiv$  sensor spectral response

- 5 The constant of proportionality in (1) depends on geometry and camera characteristics, but is basically the same for all pixels in the frame.

The ratio of video signals for two adjacent pixels is:

$$\frac{g_1}{g_2} = \frac{\int E_1(\lambda) r_1(\lambda) S(\lambda) d\lambda}{\int E_2(\lambda) r_2(\lambda) S(\lambda) d\lambda} = \frac{\int E(\lambda) r_1(\lambda) S(\lambda) d\lambda}{\int E(\lambda) r_2(\lambda) S(\lambda) d\lambda}$$

10

(2)

where we have used Land's approximation that the scene irradiance does not vary significantly between adjacent pixels:  $E_1(\lambda) \equiv E_2(\lambda) \equiv E(\lambda)$ . Assuming that the spectral reflectances are nearly constant over the spectral response of the camera, then  $r_k(\lambda) \equiv r_k = 1, 2$  and

15

$$\frac{g_1}{g_2} = \frac{r_1 \int E(\lambda) S(\lambda) d\lambda}{r_2 \int E(\lambda) S(\lambda) d\lambda} = \frac{r_1}{r_2}$$

(3)

- 20 In other words, for the conditions specified, ratios of adjacent pixel values satisfy the requirement of being determined by scene reflectances only and are independent of scene illumination. It remains to consider the practicality of the approximations used to arrive at (3). A basic assumption in the retinex process is that of only gradual spatial variations in the scene irradiance; that is, we must have nearly the same irradiance of adjacent pixel areas in the scene. This assumption is generally true for diffuse lighting, but for  
 25 directional sources it may not be. For example, the intrusion of a light beam into the area being viewed can introduce rather sharp shadows, or change the amount of light striking a vertical surface without similarly changing the amount of light striking an adjacent tilted surface. In these instances, ratios

between pixels straddling the shadow line in the first instance, or the surfaces in the second instance, will change even though no object has been introduced into the scene. However, even in these cases, with 512 by 484 resolution, the pixel-to-pixel change is often less than it appears to the eye, and the changes only  
5 appear at the boundaries, not within the interiors of the shadows or surfaces. By establishing a threshold on hits, the system can tolerate a number of these hits without triggering an intrusion alarm.

Another method, based on edge mapping, is also possible. As in the previous situation, the edge mapping process would be employed after an initial  
10 detection stage is triggered by pixel value changes from one frame to the next. Within each detected "blob" area, an edge map is made for both the initial (unchanged) frame and the changed frame that triggered the alert. Such an edge map can be constructed by running an edge enhancement filter (such as a Sobel filter) and then thresholding. If the intrusion is just a light change, then the  
15 edges within the blob should be basically in the same place in both frames. However, if the intrusion is an object, then some edges from the initial frame will be obscured in the changed frame and some new edges, internal to the intruding object, will be introduced.

Extensive laboratory testing revealed problems with both methods. In  
20 particular, it is difficult to set effective thresholds with the retinex method, because with a background and intrusive object both containing large uniform areas, many adjacent pixel ratios of unity in both the reference frame and the new frame are obtained. Therefore the fraction of ratios that are changed is diluted by those which contribute no information one way or the other. On the  
25 other hand, the edge mapping method shows undue dependence on light changes because typical edge masks use absolute differences in pixel values. Light changes can cause new edges to appear, or old ones to disappear, in a binary edge map even through there is no intervening object. By exploiting concepts from both methods, and key to this invention, an algorithm having  
30 both good detection and false alarm performance characteristics has been

constructed. Additional system features also help eliminate light changes of certain types which are expected to occur, so to further enhance performance.

5 The basic premise of the variable light rejection algorithm used in the method of the invention is to compare ratios of adjacent pixels from a segmented area in frame f1 with ratios from corresponding pixels in frame f3, but to restrict the ratios to those across significant edges. Restricting the processing to ratios of pixels tends to reject illumination changes, and using only edge pixels eliminates the dilution of information caused by large uniform areas.

10 In implementing the algorithm,

- a) Ratios  $R$  of adjacent pixels (both horizontally and vertically) in frame f1 are tested to determine if they significantly differ from unity:  $R-1 > T_1$ ? or  $(1/R)-1 > T_1$ ?, where  $T_1$  is a predetermined threshold value. Every time such a significant edge pair is found an edge count value is incremented.
- 15 b) Those pixel pairs that pass either of the tests in a) have their corresponding ratios  $R'$  for frame f3 calculated.
- c) A check is made to see if  $R'$  differs significantly from the corresponding ratio  $R$ :  
 $|R'-R|/R > T_2$ ?, where  $T_2$  is a second predetermined threshold value. Each time  
20 this test is passed a hit count value is incremented.
- d) A test is made for new edges in frame f3 (i.e., edges not in frame f1):  $R'-1 > T_1$ ? or  $(1/R')-1 > T_1$ ? Every time such a new significant edge pair is found the edge count value is incremented again.
- e) Those pixel pairs that pass either of the tests in d) have their  
25 corresponding ratios from frame f1,  $R$ , calculated.
- f) A check is made to see if ratio  $R'$  differs significantly from the corresponding ratio  $R$ :  $|R'-R|/R > T_2$ ? Each time this test is passed the hit count value is incremented again.
- g) The segmented area is now deemed an intrusion if the ratio of changed  
30 edges to the edge count value (ecv) is sufficiently large: that is, there is an

intrusion if  $H/ecv > T_3$ , where  $T_3$  is a third predetermined threshold value.

#### SHADOW REMOVAL

While the object is being outlined and segmented, the x and y coordinates of the pixels outlined and segmented are accumulated. This information is now used to calculate the centroid Z (see Fig. 6B) of the object. Also, the minimum and maximum x and y pixel coordinates of the object are computed at this time (see Fig. 6B). Both the centroid of the object and the object's minimum and maximum x, y coordinate values are used in a process to remove a shadow S (see Fig. 6A) from the object. Using the coordinate values, and assuming that life forms exhibit compact mass shapes, pieces of the object which stick out can be identified as a shadow and can be curtailed during subsequent processing. For drawing simplification, object O is shown in Fig. 4B with its shadow S removed.

#### SHAPE FEATURES

Having outlined and region grown an object to be recognized, a series of linear shape features and Fourier descriptors are extracted for each segmented region. Values for shape features are numerically derived from the image of the object based upon the x, y pixel coordinates obtained during outlining and segmentation of the object. These features include, for example, values representing the height of the object ( $y_{max} - y_{min}$ ), its width ( $x_{max} - x_{min}$ ), horizontal and vertical edge counts, and degree of circularity. However, it will be understood that there are a large number of factors relating to the features of an object and that some, or all, of the above listed features can be used with combinations of these other factors in order to classify an object. What is important is that any feature factor selected facilitate the distinction between a human and a non-human class of objects.

#### FOURIER DESCRIPTORS

Fourier descriptors represent a set of features used to recognize a silhouette or contour of an object. As shown in Fig. 6C, the outline of an object is resampled into equally spaced points located about the edge of the object.

The Fourier descriptors are computed by treating these points as complex points and creating a point complex FFT (Fast Fourier Transform) for the sequence. The resulting coefficients are a function of the position, size, orientation, and starting point P of the outline. Using these coefficients, Fourier descriptors are  
5 extracted which are invariant to these variables. As a result of performing the feature extractions, what remains is a set of features which now describe the segmented object.

#### FEATURE SET NORMALIZATION

The feature set obtained as described above is now normalized. For  
10 example, the set of features may be rescaled if the range of values for one of the features of the object is larger or smaller than the range which the rest of the features of the object have. Further, a test data base is established and when the feature data is tested on this data base, a feature may be found to be skewed. To eliminate this skewing, a mathematical function such as a logarithmic function  
15 is applied to the feature value. To further normalize the features, each feature value may be exercised through a linear function; that is, for example, a constant value is added to the feature value, and the result is then multiplied by another constant value. It will be understood that other consistent descriptors such as wavelet coefficients and fractal dimensions can be used instead of  
20 Fourier descriptors.

#### OBJECT CLASSIFIER

Having normalized a feature set, the set is now evaluated in order to classify the object which is represented by the set. An object classifier portion of the processor means is provided as an input the normalized feature set for the  
25 object to be classified. The object classifier has already been provided feature set information for humans as well as for a variety of animals (cat, dog, bird) such as shown in Figs. 7A - 7C. These Figs. show the presence of each animal in an actual scene as viewed by the camera of the system. By evaluation the feature set for the object with those for humans and animals, the classifier can  
30 determine a confidence value for each of three classes: human, animal and

unknown. Operation of the classifier includes implementation of a linear or non-linear classifier. A linear classifier may, for example implement a Bayes technique, as is well known in the art. A non-linear classifier may employ, for example, a neural net which is also well-known in the art, or its equivalent.

5 Regardless of the object classifier used, operation of the classifier produces a "hard" decision as to whether the object is human, non-human, or unknown. Further, the method involves using the algorithm to look at a series of consecutive frames in which the object appears, perform the above described sequence of steps for each individual frame, and integrate the results of the  
10 separate classifications to further verify the result.

Depending upon the outcome of the above analysis, the processing means, in response to the results of the object classification provides an indication of an intrusion if the object is classified as a human. It does not provide any indication if the object is classified as an animal. This prevents  
15 false alarms. It will be understood, that because an image of a scene provided by a camera C is evaluated on a continual basis, every one-half second for example, the fact that a human is now present in the scene but the result of the classification process may not identify him as such at one instant, does not mean that the intrusion will be missed. Rather, it only means that the human was not  
20 recognized as such at that instant. Because the movement of a human intruder into and through the scene involves motion of the person's head, trunk, and limbs, their position or posture will be recognized as those of a human, if not in one image of the scene, then probably in the next. And, anytime the presence of a human intrusion is continually recognized in accordance with the method of  
25 the invention, the alarm is given. Moreover, if the result of an object classification is unknown, an alarm indication is also given. However, the level of the alarm is less than that for a classified human intrusion. What this lower level alarm does is to alert security personnel that something has occurred which may require investigation. This is important because while the system is  
30 designed to not provide false alarms, it is also designed to not miss any human

intrusions either. Because of the manner in which the algorithm is constructed, the possibility an object will be classified as unknown is very small. As a result, the instances in which a low level alarm will be sounded will be infrequent. This is a much different situation than sounding an alarm every time  
5 there is an anomaly.

An alarm, when it is given, is transmitted to a remote site such as a central monitoring location staffed by security personnel and from which a number of locations can be simultaneously monitored.

In view of the foregoing, it will be seen that the several objects of the  
10 invention are achieved and other advantageous results are obtained.

As various changes could be made in the above constructions without departing from the scope of the invention, it is intended that all matter contained in the above description or shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.



Claims

1. A video security system visually monitoring a scene and detecting the presence of a human intruder within the scene comprising imaging means continually viewing the scene and producing a signal representative of the scene;  
5 processor means processing the signal, comparing the signal representing the scene at one point in time with a similar signal representing the scene at a previous point in time, and identifying those segments of the scene at said one point in time which differ from segments of the scene at the earlier point in time; and, discriminator means evaluating those segments of the scene identified as  
10 being different to classify each segment as a human life form or not, to give an alarm whenever an object present in one of the segments is classified as a human life form representing a human intruder within the scene, and to give no alarm if objects present in the segments are classified as non-human life forms.
2. The video security system of claim 1 wherein said discriminator means  
15 includes means comparing pixel elements contained in each segment of the scene at one point in time and corresponding pixel elements contained in a corresponding segment from the scene at the earlier point in time, and producing an outline of each segment within the later scene which differs from a corresponding segment in the earlier scene.
- 20 3. The video security system of claim 2 wherein said discriminator means further includes means growing each segment to a size which incorporates all of the pixels which define an object contained within the segment.
4. The video security system of claim 3 wherein said discriminator means further includes means extracting a set of features from the object.
- 25 5. The video security system of claim 4 wherein said feature extraction means includes means extracting linear shape features from the object as numerical values representing such factors as the height, width, horizontal and vertical edges of the object, and degree of circularity of the object.
6. The video security system of claim 4 wherein said feature extraction  
30 means further includes means extracting Fourier descriptors of the silhouette

shape features of the object.

7. The video security system of claim 6 wherein said feature extraction means further includes means normalizing any value obtained from the feature extraction means in the event said value falls outside a predetermined range of values for the particular feature.

8. The video security system of claim 6 wherein said discriminator means further includes classifier means evaluating said set of features for said object with sets of features representing human and non-human life forms and for deriving a value representing a degree of confidence as to the correspondence of the object to a human or non-human life form.

9. The video security system of claim 8 further including means providing an alarm indication only if the degree of confidence for the correspondence of the object to a human life form exceeds a predetermined confidence level.

10. The video security system of claim 8 wherein said classifier means includes a linear object classification means providing a confidence level output for each of the three classes: human, animal, and unknown.

11. The video security system of claim 8 wherein said classifier means includes a non-linear object classification means providing a confidence level output for each of three classes: human, animal, and unknown.

12. The video security system of claim 1 wherein said discrimination means includes means executing an algorithm to perform object classification.

13. The video security system of claim 4 wherein said feature extraction means includes means eliminating shadows cast by an object represented by the segment.

14. The video security system of claim 9 wherein said alarm indication means further provides a second alarm indication if an object is classified as unknown.

15. A video security system visually monitoring a scene and detecting motion of an object within the scene comprising imaging means continually viewing the scene and producing a signal representative of the scene; processor means processing said signal, comparing the signal representing the scene at one point in

time with a similar signal representing the scene at a previous point in time, and identifying those segments of the scene at said one point in time which differ from segments of the scene at the earlier point in time; and, discriminator means evaluating those segments of the scene identified as being different to determine if

5 the differences are caused by surface differences which are indicative of the presence of an intruder within the scene, or lighting changes which occur within the scene and do not indicate the presence of an intruder, and if the difference is caused by the presence of an intruder providing an indication thereof, said discriminator means including means comparing pixel elements contained in each

10 segment of the scene at the one point in time and corresponding pixel elements contained in a corresponding segment from the scene at the earlier point in time, and means determining a ratio of light intensity between each pixel in a segment with each pixel adjacent thereto, and means comparing the ratio values for the pixels in the segment of the scene at one point in time with the ratio values for the

15 pixels in the corresponding segment of the scene at the earlier point in time.

16. A method of evaluating a scene to determine if any perceived movement within the scene is caused by an intruder into the scene comprising viewing the scene and creating an image of the scene, said image of said scene comprising a plurality of pixels arranged in an array; comparing the image of the scene with a

20 reference image thereof to produce a difference image, producing said difference image including convolving the image with an antialiasing means to eliminate any aliasing effects in the resulting difference image, outlining any segments where a possible movement has occurred, determining a ratio of light intensity between each pixel in a segment with each pixel adjacent thereto, and comparing the ratio

25 values for the pixels in a segment of one image with the ratio values for the pixels in the corresponding segment of other image; processing the difference image to identify any segments therewithin which, based upon a first predetermined set of criteria, represent spatially constrained movements of an object fixed within the scene, and further processing the difference image to identify any segments

30 therewithin which, based upon a second predetermined set of criteria, represent

artifacts not caused by the presence of an intruder within the scene, said segments meeting said first and second sets of criteria being identified as segments not requiring further processing; and, further processing those segments within the difference image which remain to determine if movement therewithin is caused by

5 an intruder.

1/ 6

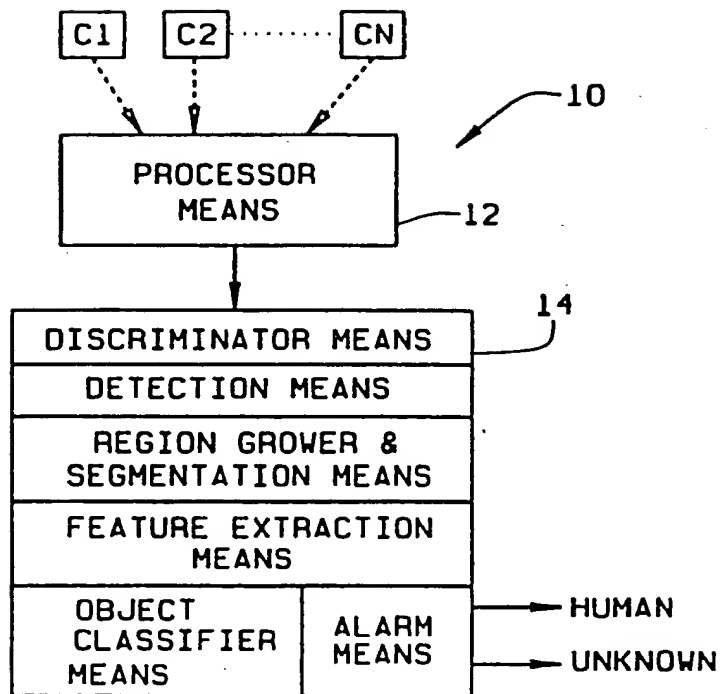


FIG. 1

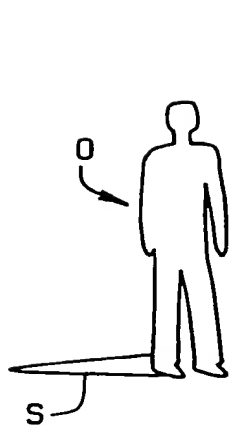


FIG. 6A

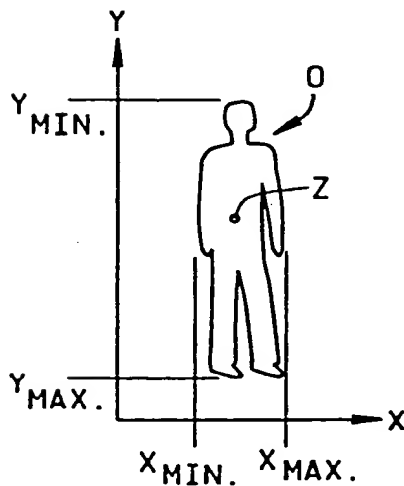


FIG. 6B

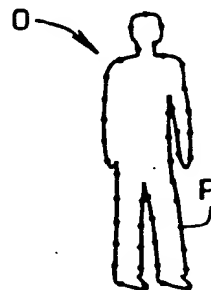


FIG. 6C

2/6

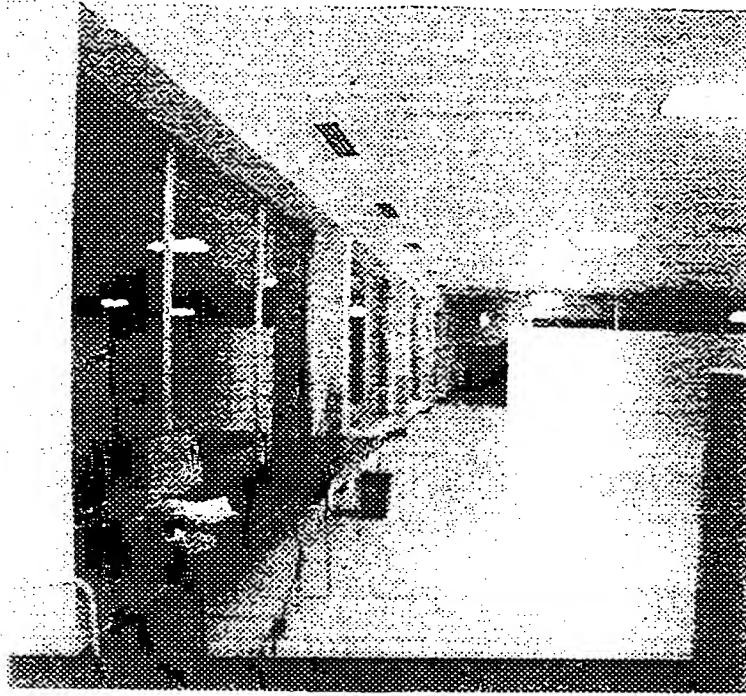


FIG. 2

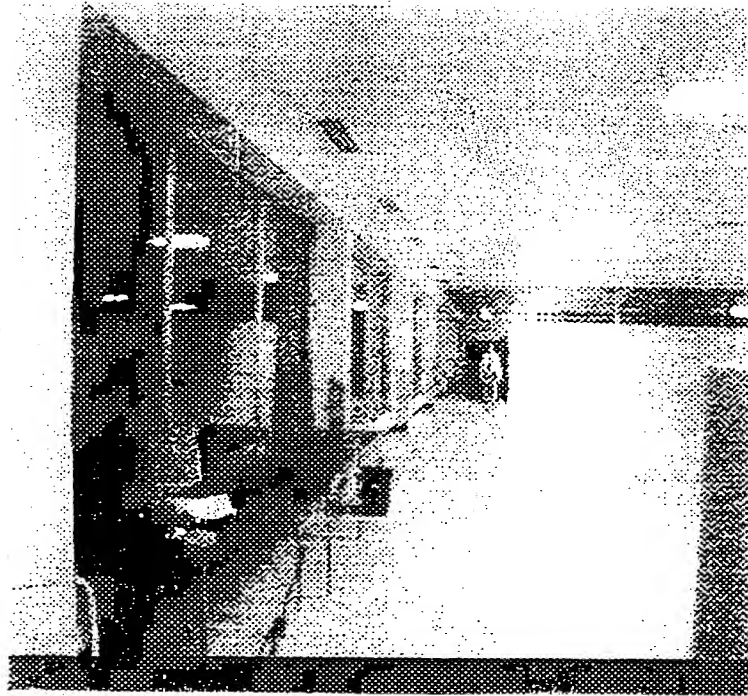


FIG. 3

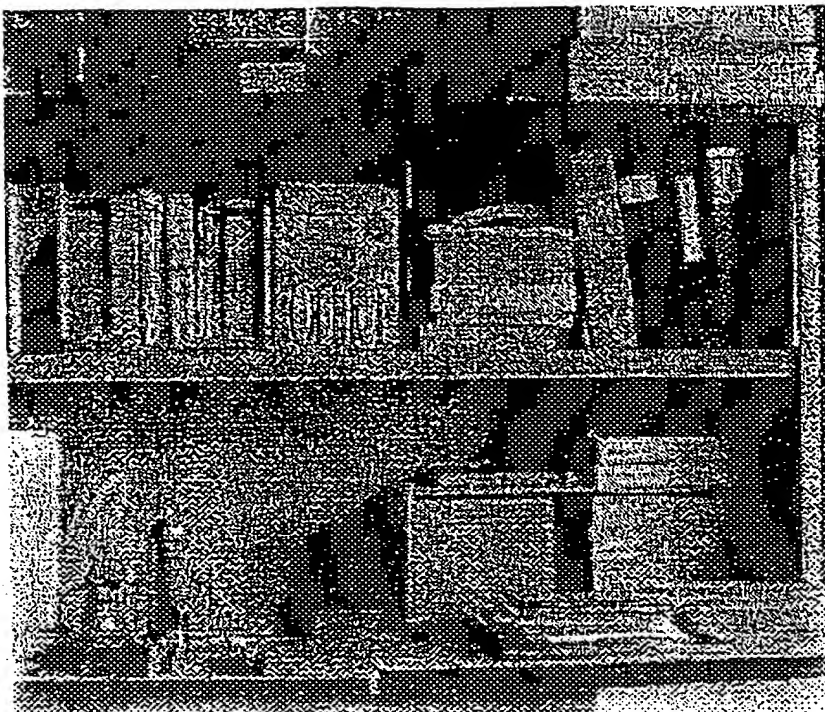


FIG. 4

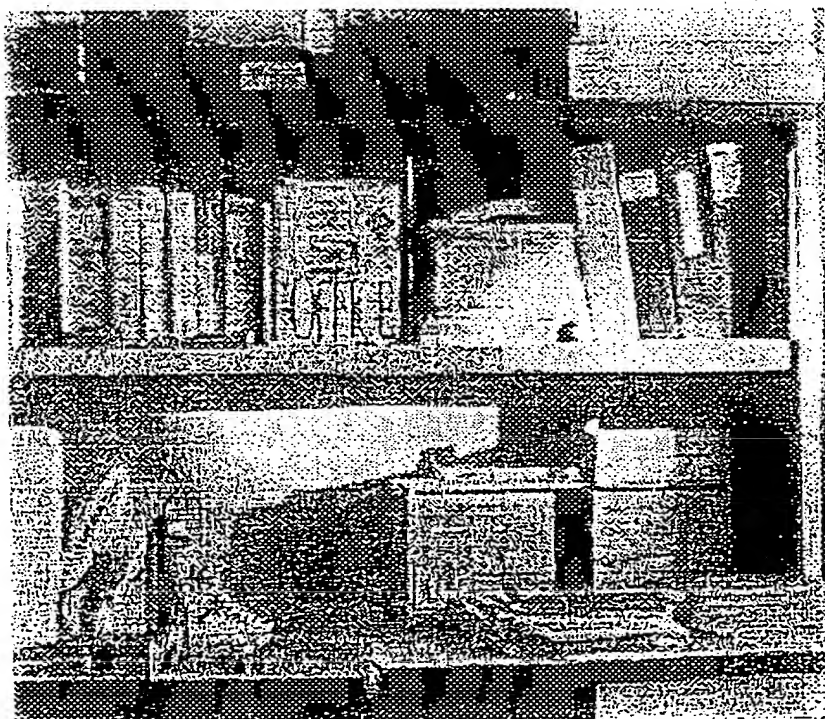


FIG. 5

4 / 6

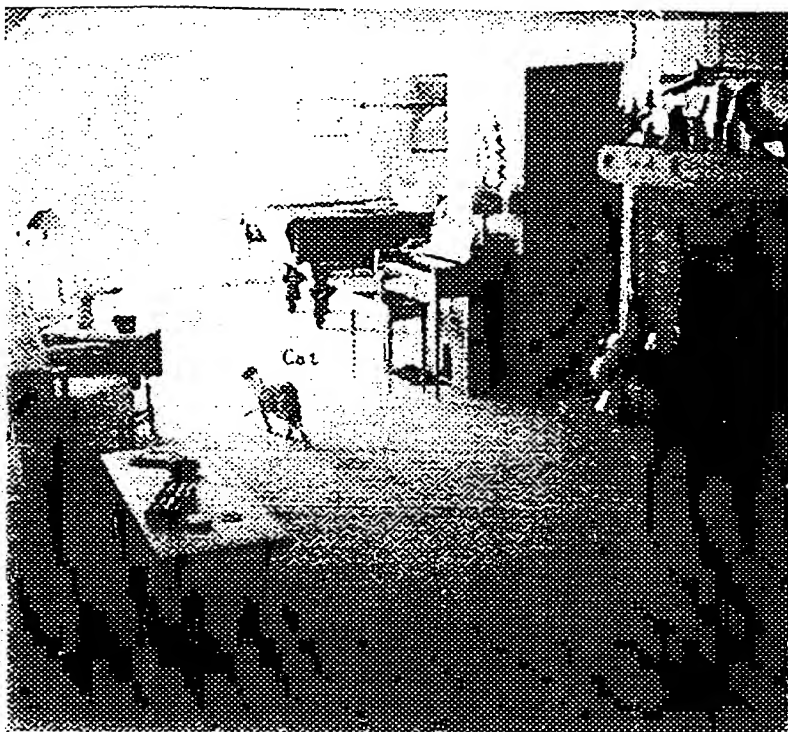


FIG. 7A



FIG. 7B



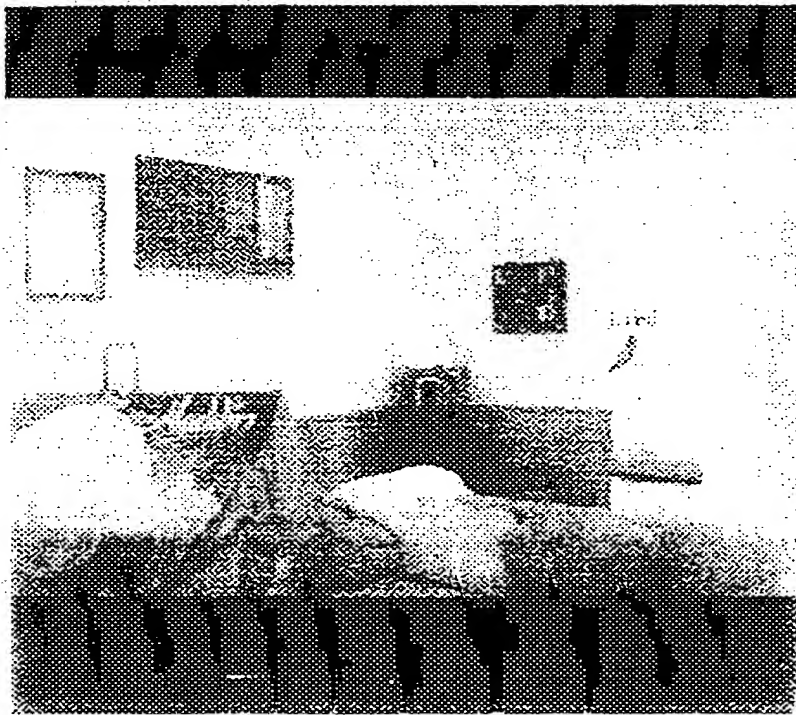


FIG. 7C

6/6

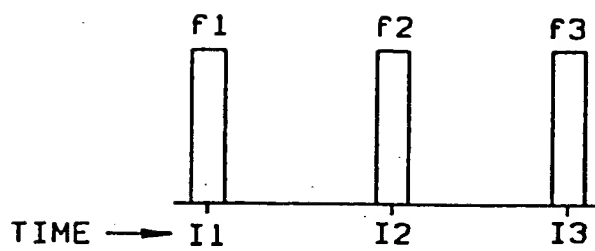


FIG. 8

P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>
P <sub>4</sub>	P	P <sub>5</sub>
P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>

FIG. 9

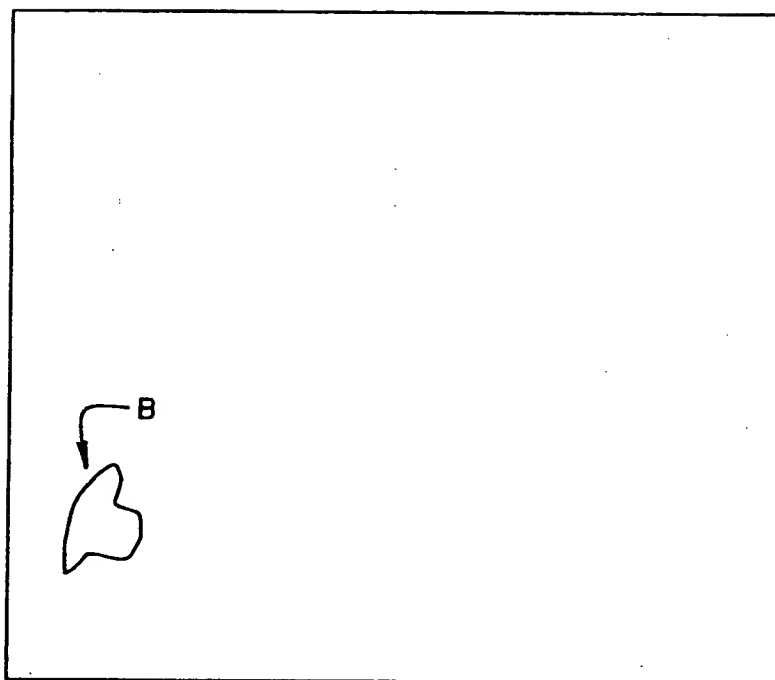


FIG. 10

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/24163

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06K 9/20, 9/46; G08B 23/00; HO4N 7/18

US CL : 340/541; 348/152, 155; 382/203, 283

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 340/541, 441, 555; 348/152, 155, 154, 143, 399, 700; 382/203, 283, 100, 192, 153; 250/342

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN

search terms: intru? and (scene# or image#), retinex, brightness(5a)change#

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Please See Continuation of Second Sheet.	

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search


23 MARCH 1998

Date of mailing of the international search report

03 JUN 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer  
  
David K. Moore

Telephone No. (703) 308-7452

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/24163

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KANETA, M. et al. "Image Processing Method for Intruder Detection around Power Line Tower," IEICE Trans. Inf. & Syst., October 1993, vol. E76-D, No. 10, pages 1153-1160.	1-5, 12-13
---		-----
Y		6-11, 14
---		-----
A	GONZALEZ R.C. et al. "Digital Image Processing," Addison-Wesley Publishing, Reading, MA, 1977, pages 320-322, 345, 348-352, 383.	15-16
Y		6-11, 14
Y		8-11, 14
Y		11
Y		14
	US 5,144,685 A (NASAR et al.) 01 September 1992, col. 10, line 59 to col. 12, line 13.	
	US 5,465,308 A (HUTCHESON et al.) 07 November 1995, col. 17, line 17 to col. 23, line 40.	
	US 5,493,273 A (SMURLO et al.) 20 February 20 1996, col. 3, lines 18-21.	